ESTIMATING TOTAL HUMAN-CAUSED MORTALITY FROM REPORTED MORTALITY USING DATA FROM RADIO-INSTRUMENTED GRIZZLY BEARS

STEVE CHERRY, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717, USA, email: cherry@math.montana.edu
MARK A. HAROLDSON, U.S. Geological Survey Northern Rocky Mountain Science Center, Interagency Grizzly Bear Study Team, Forestry Sciences Lab, Montana State University, Bozeman, MT 59717, USA, email: mark_haroldson@usgs.gov
JAMES ROBISON-COX, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717, USA, email: jimrc@math.montana.edu
CHARLES C. SCHWARTZ, U.S. Geological Survey Northern Rocky Mountain Science Center, Interagency Grizzly Bear Study Team, Forestry Sciences Lab, Montana State University, Bozeman, MT 59717, USA, email: chuck_schwartz@usgs.gov

Abstract: Tracking mortality of the Yellowstone grizzly bear (*Ursus arctos horribilis*) is an essential issue of the recovery process. Problem bears removed by agencies are well documented. Deaths of radiocollared bears are known or, in many cases, can be reliably inferred. Additionally, the public reports an unknown proportion of deaths of uncollared bears. Estimating the number of non-agency human-caused mortalities is a necessary element that must be factored into the total annual mortality. Here, we describe a method of estimating the number of such deaths from records of reported human-caused bear mortalities. We used a hierarchical Bayesian model with a non-informative prior distribution for the number of deaths. Estimates of reporting rates developed from deaths of radio-instrumented bears from 1983 to 2000 were used to develop beta prior probability distributions that the public will report a death. Twenty-seven known deaths of radio-instrumented bears occurred during this period with 16 reported. Additionally, fates of 23 radio-instrumented bears were unknown and are considered possible unreported mortalities. We describe 3 ways of using this information to specify prior distributions on the probability a death will be reported by the public. We estimated total deaths of non-instrumented bears in running 3-year periods from 1993 to 2000. Thirty-nine known deaths of non-instrumented bears were reported during this period, ranging from 0 to 7/year. Seven possible mortalities were recorded. We applied the method to both sets of mortality data. Results from this method can be combined with agency removals and deaths of collared bears to produce defensible estimates of total mortality over relevant periods and to incorporate uncertainty when evaluating mortality limits established for the Yellowstone grizzly bear population. Assumptions and limitations of this procedure are discussed.

Ursus 13:175–184 (2002)

Key words: Bayesian statistics, grizzly bear, mortality, *Ursus arctos,* Yellowstone

Management of grizzly bears requires monitoring the number of deaths occurring annually. Deaths are recorded in a number of ways: deaths of radio-instrumented individuals, deaths of animals as a result of management actions (i.e., removal of nuisance bears), and reports of deaths by the general public. However, not all bears that die are reported and an estimate of unreported deaths can improve and refine management. Estimation of the number of human-caused mortalities is particularly important. We describe and illustrate a Bayesian method of estimating the total number of human-caused deaths. The results presented here are for illustrative purposes only. Readers should not interpret these results as our final analysis on the estimation of total human-caused mortality in the Yellowstone grizzly bear population.

STUDY AREA

Our study area (centered at latitude 44.64°N, 110.52°W) contains approximately 37,500 km^2 in the states of Wyoming, Montana, and Idaho and encompasses Yellowstone National Park and portions of 6 National Forests that surround the park. A primary component of the occupied grizzly bear habitat within the area is designated as the Yellowstone Grizzly Bear recovery zone (U.S. Fish and Wildlife Service 1993). During the last decade and a half, grizzly bears in the Greater Yellowstone Ecosystem (GYE) have expanded their range (Schwartz et al. 2002), and an increasing number of mortalities are occurring outside of the designated recovery zone. We included all relevant bear mortalities from the GYE in our analysis without regard to their specific location. Detailed descriptions of the study area can be found in Knight and Eberhardt (1985), Blanchard and Knight (1991), and Mattson et al. (1991).

METHODS

Determination of Death

Data on grizzly bear mortalities were obtained from the Montana Department of Fish, Wildlife and Parks, which maintains the official database on grizzly bear mortalities in the GYE. We focused our analysis on 1983–2000, after the Interagency Grizzly Bear Committee (IGBC) was formed. The IGBC implemented regulations designed to minimize human-caused grizzly bear mortalities (IGBC 1986). Our analysis uses the number of publicly reported or known human-caused grizzly bear deaths to estimate the total number of human-caused mortalities. We excluded (1) all natural bear deaths, (2) agency-sanctioned management removals, (3) mortalities of radio-instrumented grizzly bears, and (4) known mortalities (i.e., carcass in hand) of bears whose cause of death was undetermined. There were 12 natural bear deaths (8 adults) and 8 bears in the undetermined category during 1983–

2000.

The degree of certainty associated with each record in the mortality database is classified as: (1) known, where carcass was recovered or other evidence to indicate known status was available; (2) probable, with strong evidence to indicate a mortality had occurred but no carcass was recovered; and (3) possible, with presumptive evidence of a mortality but no prospects for validation (Craighead et al. 1988). We used known and probable human-caused mortalities for analysis under one scenario and included possible human-caused mortalities for a second analysis that is more inclusive of all possible human-caused grizzly bear deaths.

Information used to estimate the percent of mortalities that are reported by the public was obtained from Interagency Grizzly Bear Study Team (IGBST) databases for 1983–2000. The IGBST has been capturing, instrumenting, and monitoring grizzly bears within the GYE since 1975. All grizzly bears except dependent offspring (cubs or yearlings) captured during research trapping efforts were radio-instrumented. All grizzly bears involved in nuisance activity within the GYE and captured by state wildlife authorities (Wyoming, Idaho, Montana) were radio-instrumented (again with the exception of dependent offspring) and data from these individuals were included in IGBST databases. Adult bears were usually instrumented with radiocollars (Telonics, Mesa, Arizona, USA) that had breakaway canvas inserts. Independent subadult bears were instrumented with expandable collars (Blanchard 1985) or glue-on-hair transmitters. All radiotransmitters had motion sensors that reduced pulse rates if transmitters were stationary for a specified period of time, usually 4–5 hours. When pulse rates and locations from aerial telemetry indicated a stationary signal over several flights (usually a minimum of 2 weeks), a field crew investigated.

Stationary signals were usually cast-off transmitters. When mortalities or collars found under suspicious circumstances were discovered, the appropriate law enforcement agencies were notified to investigate the cause of death. Instances in which transmitters could not be retrieved and the individual was never recaptured were designated "unresolved losses". We believed there was a strong probability that some unknown portion of these instances were human-caused mortalities. Transmitters located in logjams in rivers or in cliffs may have been purposefully discarded in these locations.

We classified radio-instrumented bears as "unexplained losses" if premature failure of a working transmitter occurred that was not logically attributed to the expected battery life of the transmitter and the individual was never recaptured. Once again we know that some of these bears were killed illegally and their collars destroyed. Law enforcement cases have been prosecuted in which persons had killed bears, destroyed collars, and confessed years later.

For confirmed human-caused mortalities of bears that were radio-instrumented at the time of loss, we determined whether discovery of the mortality was due to the transmitter. These mortalities were classified as unreported, discovery due to telemetry, if the public did not report them. We included individuals in this group whose radiocollars had been cut off and separated from presumably dead bears. Forensic techniques were used to determine with a high degree of certainty whether suspected cut collars had been cut. Unexplained and unresolved losses indicated by telemetry were also classified as unreported, discovery due to telemetry. Mortalities of instrumented bears that were discovered without the aid of telemetry were considered reported (i.e., public finds or discovered without telemetry).

Our classification of method of discovery for known dead bears and unexplained and unresolved loss allowed us to produce 2 estimates of the percent of human-caused mortalities that are reported by the public. The first uses only the method of discovery for confirmed human-caused mortalities. The second is considered inclusive of all known sources of possible human-caused mortality and includes unexplained and unresolved loss of radio-instrumented bears. These 2 classifications likely bound reality.

The classification of mortalities and discovery criteria we employed allowed us to produce estimates of unreported human-caused grizzly bear mortality that include or consider all possible sources of human-caused mortality (Table 1, Fig. 1). We believe this is a reasonable approach given that we are dealing with a threatened species characterized by low reproductive potential.

The Bayesian Method

Bayesian statistics is fundamentally different from the frequentist statistics taught in most introductory statistical methods courses. In a Bayesian analysis a prior (pre-data) probability distribution is used to describe uncertainty about an unknown parameter or parameters. Data are collected and used to update or modify the prior distribution to obtain the posterior (post-data) distribution for the unknown parameter. To avoid subjectivity in prior specification, some practitioners use "non-informative" priors, which often give results similar to the frequentist approach.

The posterior distribution is the ultimate goal of all Bayesian analyses. It contains all relevant information about the unknown parameter or parameters of interest. Different numerical summaries of the posterior distribution may be of interest in different settings. Common

Table 1. Definitions of terms related to cause, certainty, and discovery of grizzly bears mortalities.

Terms		Definition
Cause of mortality	Natural	Death could be positively or reasonably attributed to natural causes.
	Human-caused	Death could be positively or reasonably attributed to humans.
	Undetermined	Cause of death could not be determined.
Certainty of mortality	Known mortality	A carcass or other evidence (i.e., parts) to substantiate death.
	Probable mortality	Strong evidence to indicate mortality but no carcass recovered. Includes cases where blood, hair, or other tissues clearly indicate severe wounding, and observations suggest the bear displayed abnormal behavior.
	Possible mortality	Some presumptive evidence of a mortality but no prospects for validation. Includes defense of life situations where shots were fired but no evidence of significant wounding was found and hearsay evidence of poaching or malicious death.
	Unresolved loss	Pulse rate and stationary location of a transmitter indicated a cast-off collar or mortality, and transmitters could not be retrieved due location (i.e., cliff, log-jam in river) or failure; bear never recaptured so fate was unresolved.
	Unexplained loss	Premature failure of a working transmitter occurred that could not logically be attributed to expected battery life; bear never recaptured so loss was unexplained.
Discovery of mortality	Reported	Mortality of an instrumented or noninstrumented bear discovered without the aid of telemetry.
	Unreported	Mortality of an instrumented bear discovered due to telemetry and not reported by the public. Unexplained and unresolved losses of telemetered bears were classified as unreported under one scenario used to tally deaths.

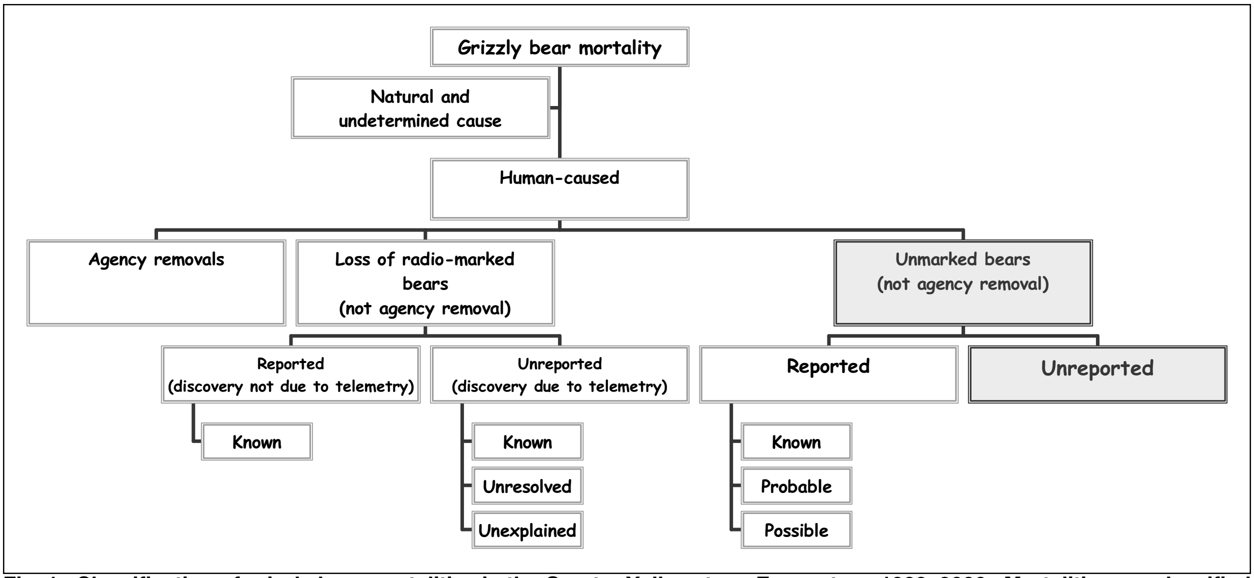

Fig. 1. Classification of grizzly bear mortalities in the Greater Yellowstone Ecosystem, 1983–2000. Mortalities are classified by type and by certainty of actual death. Estimation of unreported, unmarked mortalities (shaded boxes) was the objective of this study.

choices include the posterior median or the posterior mode. Another frequently used summary is a $100(1 - \alpha)\%$ posterior probability interval or credible interval. The Bayesian interpretation is that the fixed interval has a $100(1 - \alpha)\%$ probability of containing the random parameter. One common method of constructing credible intervals is to choose the relevant percentiles. For example, a 95% credible interval can be constructed by choosing the lower endpoint to be the 2.5^{th} and the upper endpoint to be the 97.5^{th} percentile of the posterior distribution. A more complete summary of the posterior for a discrete parameter is a graph of the cumulative posterior probability distribution (or a table of relevant percentiles when the posterior is complex). The numerical summaries can be obtained by simulating data from the posterior distribution (Gelman et al. 1995).

Ellison (1996) gives an accessible introduction to the general topic of Bayesian statistics in ecology. Dennis (1996) discusses the negative aspects of Bayesian statistics. Hilborn and Mangel (1997) discuss the use of Bayesian statistics in ecology. Lee (1997) and Box and Tiao (1973) are introductory statistical textbooks on the topic of Bayesian analysis. More technical presentations can be found in Gelman et al. (1995) and Carlin and Louis (1996). Further details of the Bayesian approach and an example can be found in Appendix I.

Consider the following scenario: We are told a coin has been tossed an unknown number of times (n) and we observe x heads. We want to estimate the number of times the coin was tossed. We do not know the probability θ of getting a head but we may be able to estimate it using a different coin that we hope is similar to the one actually tossed. In the practical setting of concern to us, n is the number of non-agency, non-collared human-caused bear deaths, x is the number of such deaths reported by the public and θ is the probability that a death will be reported.

Clearly, this is a difficult problem. It has a long history in statistics (Raftery 1988 and references therein) and is not readily solved by non-Bayesian methods. Raftery (1988) suggested a Bayesian approach for problems of this type. He started with a collection of success counts from a binomial distribution with unknown parameters n (the number of trials) and θ (the probability of a success on any one trial). A joint prior distribution is required for n and θ. Specification of prior distributions for a discrete parameter is difficult (Gelman et al. 1995), but Raftery (1988) rather cleverly solved the problem with a hierarchical Bayesian approach using

$$p(n,\theta) = p(n)p(\theta) \propto n^{-1}$$

The 2 parameters are assumed to be independent. The implied prior on θ is a uniform distribution on (0,1). That is, prior to data collection θ is assumed to be no more likely to fall into any particular interval of (0,1). This is an example of a non-informative prior on θ. Non-informative priors are used when there is little prior information available about a parameter or parameters of interest. Additional discussion of non-informative priors can be found in the Appendix. The prior distribution on n is also considered to be non-informative.

In our case, prior information on θ does exist and we have modified Raftery's (1988) method to account for this fact. The prior information comes from the radio-instrumented sample of dead bears. A common choice for a prior distribution for a probability is the beta distribution. This is a continuous distribution defined over the interval (0,1). The family of beta distributions is indexed by 2 shape parameters $a > 0$ and $b > 0$. The mean of the beta distribution is

$$\frac{a}{a+b}$$

and the variance is

$$\frac{ab}{(a+b)^2(a+b+1)}$$

The 2 shape parameters make this a flexible family of prior distributions. It will be assumed below that a and b are integers as this leads to an expression for the posterior in terms of factorial quantities.

Given these priors the joint posterior distribution of n and θ is (for $n > x$)

$$p(n,\theta \mid x) = \theta^{x+a-1}(1-\theta)^{n-x+b-1}\frac{(n-1)!(a+b-1)!}{(x-1)!(n-x)!(a-1)!(b-1)!}$$

where n, θ, and x are as defined above. The marginal posterior distribution for n is obtained by integrating θ out of the joint posterior and is given by

$$p(n \mid x) = \frac{(n-1)!(x+a-1)!(n-x+b-1)!(a+b-1)!}{(x-1)!(n-x)!(a-1)!(b-1)!(n+a+b-1)!}$$

This is a beta-negative binomial distribution. Similar results hold if a and b are any positive real numbers.

It is necessary to choose suitable values for a and b. A purely Bayesian approach would rely on prior (pre-data) knowledge of the investigator. In Raftery's (1988) original formulation of the problem, he assumed that no information was available about the probability of a success and he chose the uniform distribution on the interval (0,1) as his prior. This is equivalent to a beta distribution with $a = b = 1$. We have other information available from the fates of radio-instrumented bears.

Assuming that the reported number of deaths of instrumented bears is binomial with parameters m and θ, we could estimate θ using standard methodology (i.e., confidence intervals). In keeping with the Bayesian theme, however, we decided to take a Bayesian approach. Assuming a binomial likelihood and a non-informative beta prior ($a = b = 1$) for θ, the posterior for θ has a beta distribution with parameters $x + 1$ and $n - x + 1$ (Gelman et al. 1995:28–31, also see Appendix I). We then used this posterior distribution for the reporting rate of deaths from radiocollared bears as the prior distribution for the reporting rate of deaths in estimating the total number of non-agency non-collared human-caused deaths of bears. We used data from 1983–2000 to estimate the priors for θ.

Including unexplained and unresolved mortalities yielded a sample size for this period of $n = 50$ with $x = 16$ reported deaths yielding a beta ($a = 17$, $b = 35$) prior with mean

$$\frac{17}{52} = 0.327$$

and standard deviation 0.0066. The prior density is approximately symmetrical about the mean of 0.327. Approximately 95% of the probability is distributed between 0.20 and 0.46. With unresolved and unexplained losses excluded, the sample size was 27, with 16 reported deaths. The resulting beta ($a = 17$, $b = 12$) prior for the reporting rate of non-agency, non-collared bear deaths has a mean of 0.586 and standard deviation 0.0093. Again the prior density is approximately symmetrical about the mean, with the prior probability that θ lies between 0.41 and 0.77 being approximately 95%. These 95% probability intervals are similar to the large sample 95% confidence intervals that could be constructed using standard statistical methodology.

These 2 choices for a prior distribution for θ provide reasonable bounds given the data from instrumented bears. In essence either we have given the unresolved or unexplained loss of collared bears as much weight as known mortalities or given them a weight of 0. Another reasonable approach would be to include the unresolved or unexplained loss bears but to downweight them. There were 23 unresolved or unexplained losses of signal incidents. For illustrative purposes we chose a third prior by assuming that there were 16 reported deaths with 23 unreported. In other words we assumed that 12 of the 23 unresolved or unexplained incidents were in fact deaths. This resulted in a beta ($a = 17$, $b = 24$) prior distribution for the reporting rate of non-agency non-collared human-caused deaths with a mean of 0.410 and standard deviation 0.00779.

For convenience we refer to the 3 prior distributions on the reporting rate as Prior 1 ($a = 17$, $b = 35$), Prior 2 ($a = 17$, $b = 12$), and Prior 3 ($a = 17$, $b = 24$). Relevant assumptions are the following:

1. Deaths occur independently of one another.
2. The probability that the death of a radio-instrumented bear is reported by the public is approximately equal to the probability that the death of a non-instrumented bear is reported by the public.
3. The probability that a death is reported is independent of the cause of death.
4. The probability that a death is reported is constant from 1983 to 2000.

We applied the method to reported bear mortalities from 1993 to 2000 in 3-year running blocks. We applied the method with possible mortalities included and excluded. We used the 3 prior distributions discussed above. Thus, there were 6 possible data and prior distribution combinations. All calculations were done in the SPLUS programming language (MathSoft Inc., Seattle, Washington, USA).

RESULTS

The data analyzed below are reported human-caused known, probable, and possible deaths of non-instrumented bears in the GYE from 1993 to 2000 (Table 2), excluding agency removals. Four deaths in 1997 were from one incident (an adult female and 3 yearlings). These do not represent 4 independent deaths and they were pooled into one death for that year.

The posterior and prior distributions for the reporting rate θ belong to the same family. We chose to summarize the results for n by reporting the cumulative posterior probability distribution for n (Figs. 2, possibles included and 3, possibles excluded) for the six 3-year periods. The posterior distribution for the number of deaths is sensitive both to the specified prior distribution on the reporting rate and to the decision of whether to include possible mortalities as deaths (Tables 3, 4).

Table 2. Reported non-instrumented grizzly bear deaths in the Greater Yellowstone Ecosystem by year and running 3-year totals, 1993–2000. Excluded are (1) all natural bear deaths, (2) agency-sanctioned management removals, (3) mortalities of radio-instrumented grizzly bears, and (4) known mortalities (i.e., carcass in hand) of bears whose cause of death was undetermined.

	Possibles included		Possibles excluded	
Year	Annual	3-year sum	Annual	3-year sum
1993	1		1	
1994	4		3	
1995	7	12	7	11
1996	4	15	4	14
1997	6	17	5	16
1998	1	11	1	10
1999	7	14	5	11
2000	16	24	13	19
Total	46		39	

The interpretation of Tables 3 and 4 is straightforward. For example, when possible deaths are included and the reporting rate is assigned Prior 1 the probability that 36 or fewer bears died during 1993–1995 is 50% and there is a 95% probability that the number of deaths was between 20 and 69. If Prior 2 is assigned, the probability that 20 or fewer bears died is 50% and there is a 95% probability that between 14 and 34 bears died. Other summary measures may be more appropriate for particular applications. Specified percentiles (e.g. quartiles) of the distribution could be determined. If a threshold value of a certain number of mortalities is specified, then one can determine the probability of having exceeded that threshold.

DISCUSSION

The Bayesian analysis is sensitive to the choice of the prior distribution for θ. Lower reporting rates result in higher estimates of total mortality and in a more widely dispersed posterior probability distribution for n. Berger et al. (1999) also documented this dependency on the prior. It is clear that an appropriate prior distribution will need to be specified for θ and careful thought is needed concerning the form of that prior. Ideally, scientists using this method would reach consensus on the form of the prior.

The priors we chose for θ used information from radio-instrumented bears. We assumed that the reporting rate for non-instrumented bears would be similar to the reporting rate for non-collared bears, and this assumption has been questioned. For example, if the presence of a collar serves as an incentive for reporting a mortality, then

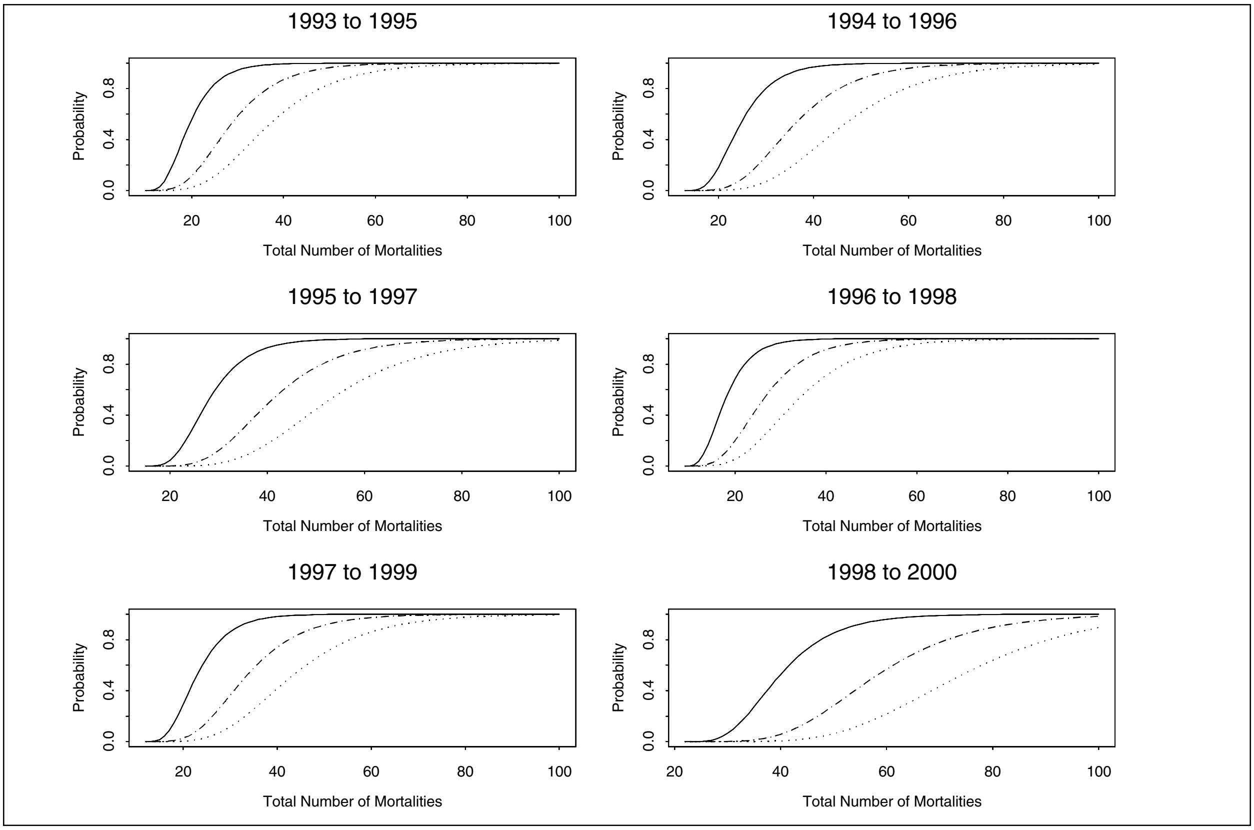

Fig. 2. Cumulative posterior distributions for the total number of human-caused mortalities (possibles included) of grizzly bears in the Greater Yellowstone Ecosystem, 1993–2000. Results are for 3 prior distributions for the reporting rate (solid line $a = 17$, $b = 12$; dash-dot $a = 17$, $b = 24$; dashed line $a = 17$, $b = 35$).

the reporting rate for instrumented bears will be higher than for non-instrumented bears. The reporting rate for instrumented bears could be less than that for non-instrumented bears if, for example, managers remove instrumented bears once it is known they are dead. This would reduce the time these carcasses are available for discovery. In a sense, the 3 prior distributions we used for the reporting rate bracket these 2 possibilities and show the potential effect of such biases.

Another critical assumption is that bear deaths are reported independently of one another. In general this assumption seems reasonable for the GYE but it can be violated as it was in 1997 when 4 bears (a female and her 3 yearlings) died in a single incident. It may not be reasonable for other geographic areas. Useful results may still be possible if incidents in which bears die can be considered independent of one another and if reliable estimates of the number of deaths per incident can be calculated.

The assumption of a constant reporting rate for radiocollared bears over time was important in our specification of the prior and in application of the method to the count data. This assumption could be violated if, for example, the probability of a death being reported depended on the cause of death and these causes changed over time. Mattson (1998) argued that this has in fact occurred. We did not have sufficient data to statistically examine the time series of radio-instrumented deaths to see if there was a trend over time, although we attempted to do so. However, there is some evidence that reporting rates have declined in recent years.

It is also apparent that decisions concerning what constitutes a dead bear are necessary both in using instrumented bears to aid in specification of the prior for the reporting rate and in choosing the sample to which the method will be applied. As indicated above, we did not include known mortalities of bears whose cause of death was undetermined. Including those deaths in the analysis, (i.e., assuming that those deaths were human-caused) would lead to a more conservative approach. However, given that very few adult bears are known to have died from causes other than humans since 1983, an alternative approach is to estimate total mortality. This would require revising current conservation strategies to focus on

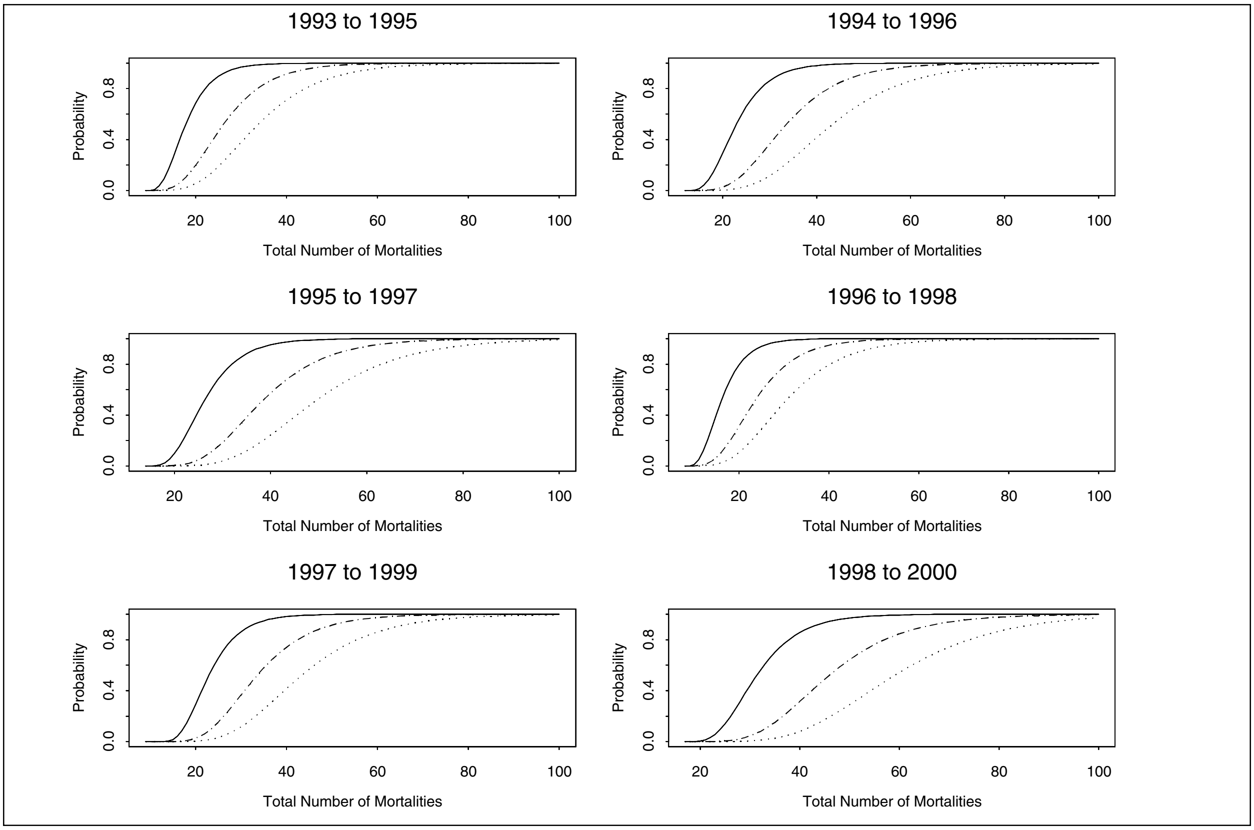

Fig. 3. Cumulative posterior distribution for the total number of human-caused mortalities (possibles excluded) of grizzly bears in the Greater Yellowstone Ecosystem, 1993–2000. Results are for 3 prior distributions for the reporting rate (solid line *a* = 17, *b* = 12; dash-dot *a* = 17, *b* = 24; dashed line *a* = 17, *b* = 35).

Table 3. Medians and 95% posterior probability intervals (in parentheses) for the posterior distribution of the total number of human caused mortalities (possibles included) for grizzly bears in the Greater Yellowstone Ecosystem, 1993–2000. Results are shown for 3 beta prior distributions on the reporting rate. Beta distribution shape parameters are indicated in column headings.

Years	Prior 1 $a = 17, b = 35$	Prior 2 $a = 17, b = 12$	Prior 3 $a = 17, b = 24$
1993–95	36 (20, 69)	20 (14, 34)	28 (17, 52)
1994–96	45 (27, 83)	25 (17, 42)	36 (22, 64)
1995–97	52 (31, 93)	29 (20, 47)	41 (25, 71)
1996–98	33 (18, 64)	18 (12, 32)	26 (15, 48)
1997–99	42 (25, 78)	23 (16, 39)	33 (20, 60)
1999–2000	73 (46, 127)	40 (29, 64)	57 (37, 97)

Table 4. Medians and 95% posterior probability intervals (in parentheses) for the posterior distribution of the total number of human caused mortalities (possibles excluded) for grizzly bears in the Greater Yellowstone Ecosystem, 1993–2000. Results are shown for 3 beta prior distributions on the reporting rate. Beta distribution shape parameters are indicated in column headings.

Years	Prior 1 $a = 17, b = 35$	Prior 2 $a = 17, b = 12$	Prior 3 $a = 17, b = 24$
1993–95	33 (18, 64)	18 (12, 32)	26 (15, 48)
1994–96	42 (25, 78)	23 (16, 39)	33 (20, 60)
1995–97	48 (29, 88)	27 (19, 44)	38 (24, 67)
1996–98	30 (16,59)	17 (11, 29)	24 (14, 45)
1997–99			
1999–2000	33 (18, 64)	18 (12, 32)	26 (15, 48)
	58 (35, 103)	32 (22, 52)	45 (29, 79)

total, rather than human-caused, mortality.

ACKNOWLEDGMENTS

We acknowledge the helpful comments of R. Boik, L. Eberhardt, D. Mattson, L. Metzgar, C. Servheen, J. Waller, G. White, and members of the Interagency Grizzly Bear Study Team. Acknowledgment does not imply agreement with our opinions on the usefulness of the method.

LITERATURE CITED

BERGER, J.O., B. LISEO, AND R.L. WOLPERT. 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 14:1–28.

BLANCHARD, B. 1985. Field techniques used in the study of grizzly bears. Interagency Grizzly Bear Study Team Report, Forestry Sciences Labratory, Montana State University, Bozeman, Montana, USA.

BLANCHARD, B.M., AND R.R. KNIGHT. 1991. Movements of Yellowstone grizzly bears, 1975–87. Biological

Conservation 58:41–67.
BOX, G.E.P, AND G.C. TIAO. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Massachusetts, USA.
CARLIN, B.P., AND T.A. LOUIS. 1996. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, New York, New York, USA.
CRAIGHEAD, J.J., K.R. GREER, R.R. KNIGHT, AND H.I. PAC. 1988. Grizzly bear mortalities in the Yellowstone Ecosystem, 1959–1987. Joint report of the Montana Department of Fish, Wildlife and Parks, Craighead Wildlife Institute, Interagency Grizzly Bear Study Team, and National Fish and Wildlife Foundation. Montana Department of Fish, Wildlife and Parks, Region 3, Bozeman, Montana, USA.
DENNIS, B. 1996. Discussion: Should ecologists become Bayesians? Ecological Applications 6:1095–1103.
ELLISON, A.M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6:1036–1046.
GELMAN, A., J.B. CARLIN, H.S. STERN, AND D.B. RUBIN. 1995. Bayesian data analysis. Chapman & Hall, New York, New York, USA.
HILBORN, R., AND M. MANGEL. 1997. The ecological detective: Confronting models with data. Monographs in Population Biology 28. Princeton University Press, Princeton, New Jersey, USA.
INTERAGENCY GRIZZLY BEAR COMMITTEE. 1986. Interagency grizzly bear guidelines. U.S. Forest Service, Washington, D.C., USA.
KNIGHT, R.R., AND L.L. EBERHARDT. 1985. Population dynamics of Yellowstone grizzly bears. Ecology 66:323–334.
LEE, P.M. 1997. Bayesian statistics: An introduction. Second edition. Arnold Press, New York, New York, USA.
MATTSON, D.J. 1998. Changes in mortality of Yellowstone's grizzly bears. Ursus 10:129–138.
———, B.M. BLANCHARD, AND R.R. KNIGHT. 1991. Food habits of Yellowstone grizzly bears, 1977–1987. Canadian Journal of Zoology. 69:1619–1629.
RAFTERY, A. 1988. Inference for the binomial N parameter: A hierarchical Bayes approach. Biometrika 75:223–228.
SCHWARTZ, C.C., M.A. HAROLDSON, K. GUNTHER, AND D. MOODY. 2002. Current distribution of grizzly bears in the Greater Yellowstone Ecosystem 1990–2000. Ursus 13:203–212.
U.S. FISH AND WILDLIFE SERVICE. 1993. Grizzly bear recovery plan. U.S. Fish and Wildlife Service, Missoula, Montana, USA.

APPENDIX I

A Brief Primer of Bayesian Statistics

A typical statistics methods course treats data as realizations of a random experiment (e.g., random sampling from a population). The purpose of data collection is to use the information in the data to draw some inference about unknown parameters (numerical characteristics of the population). The 2 most common forms of inference are estimation (point and interval) and testing. Interpretation of the results of such inference procedures requires 2 key assumptions. The first is that the probability of some event of interest is the proportion of times that event occurs if the basic chance process (random sampling) is repeated over and over independently and under the same conditions. The second is considering the parameter to be an unknown constant. Given the long run relative frequency interpretation of probability and the assumption of a constant but unknown parameter, it makes no sense to talk about the probability that the parameter will have some specified value or fall into some interval of values. It is a fixed constant that does not vary with repeated sampling, and it either has the value or falls in the interval or it does not. There is no probability statement to be made about the parameter. Probability models apply to data, not to parameters. These methods are thus considered a part of frequentist statistical methodology.

Bayesian statisticians reject this line of reasoning. Bayesians consider probability to be the natural language of uncertainty. They talk about the probability that a parameter will fall into some interval of values even if the parameter truly is fixed (e.g., the average weight of adult male grizzly bears in Montana on any given day). Probability to a Bayesian is a subjective assessment of the strength of a personal belief about an unknown parameter. Inference begins with quantifying the degree of prior (pre-data) uncertainty about the parameter by choosing an appropriate probability distribution for the parameter called the prior distribution (the prior). The prior is a quantitative description of what an investigator believes based on previous experience and knowledge. Data are collected and information in the data is used to update or modify the prior beliefs resulting in a posterior distribution (the posterior).

An example will help to illustrate these approaches. A common problem in introductory statistics courses involves estimation of a population proportion. The data collection procedure is conceptualized as drawing n independent observations $(X_1, X_2, \ldots, X_n)$ from a Bernoulli distribution and counting the number of successes. Each of the random variables X_i equals 0 or 1 depending on whether the observation was a failure or a success, respectively. The number of successes

$$Y = \sum_{i=1}^{n} X_i$$

is a binomial random variable with parameters n (which is known) and θ (which is unknown). The probability of a success on any one observation (θ) is the unknown parameter of interest. Although unknown, θ is considered to be a constant. The probability distribution of Y is given by

$$p(y \mid \theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

for $y = 0,1,2,\cdots,n$. The unknown parameter is estimated by computing the sample proportion $\hat{\theta} = Y/n$. The construction of confidence intervals and hypothesis tests follows in the usual way. For example, a large sample 95% confidence interval can be calculated from

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

A Bayesian analysis starts with a mathematical description of an investigator's *a priori* belief about θ. These beliefs are expressed in terms of the prior denoted here by $p(\theta)$. Suppose that we have n observations $(X_1, X_2, \ldots, X_n)$ from a Bernoulli distribution. The probability distribution of Y is a function of the data given θ. Viewed as a function of θ given the data, such a function is referred to as the likelihood, which we denote as $l(\theta \mid y)$. Although they look like identical functions, p and l are quite different because in the first case the parameter is fixed and the data vary, and in the second case the data are fixed and the parameter varies. That is

$$p(y \mid \theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \text{; for } y = 0,1,2, \ldots, n$$

and

$$l(\theta \mid y) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \text{; for } 0 < \theta < 1$$

The likelihood is used to update prior beliefs about θ quantified in $p(\theta)$ using Bayes Theorem to compute a posterior distribution $p(\theta \mid y)$ for the parameter. Bayes Theorem tells us that

$$p(\theta \mid y) = \frac{l(\theta \mid y)\, p(\theta)}{m(y)}$$

The quantity in the denominator $m(y)$ is referred to as the marginal distribution of the data, but often it is not necessary to compute this in practice. It contains no information about θ and is in fact a constant because the computation of the posterior is a post-data operation. It is enough to think of Bayes Theorem as

$$\text{posterior} \propto (\text{likelihood})\ x\ (\text{prior})$$

The posterior distribution is the goal of a Bayesian analysis. It summarizes an investigator's knowledge of the parameter given prior belief and subsequent data.

A common choice for a prior for θ is a beta distribution. The beta family of distributions is a flexible 2-parameter (a and b) family of continuous distributions defined over the interval from 0 to 1. The parameters a and b are positive real numbers. This makes it an ideal source of potential priors for a probability such as θ. It is enough for our purposes to note that with a beta prior,

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \text{, for } 0 < \theta < 1$$

Thus, the posterior is (ignoring constant terms that are independent of θ)

$$p(\theta \mid y) \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{y}(1-\theta)^{n-y} = \theta^{a+y-1}(1-\theta)^{b+n-y-1}$$

The posterior is also a beta distribution with parameters $a + y$ and $b + n - y$.

For a specific example, suppose that we choose a beta prior for θ with parameters $a = 3$ and $b = 7$. The mean of a beta random variable with parameters a and b is a/(a+b), and the variance is $(ab)/[(a + b)^2 (a + b + 1)]$. The mean for the prior given here is 0.3 and the standard deviation is 0.197. By specifying this prior we are stating our belief that the true probability of a success is around 0.3, but there is some uncertainty associated with that belief. We take a sample of size 5 and observe 4 successes. The posterior distribution is a beta distribution with parameters $a + y = 3 + 4 = 7$ and $b + n - y = 7 + 5 - 4 = 8$. The posterior mean is 0.467 and the posterior standard deviation is 0.132. Observing 4 successes in 5 trials has shifted the prior distribution to the right, giving more credence to higher values of θ. Note that the frequentist estimate of θ would be θ = 0.8.

A 100(1 – α)% posterior probability interval or credible interval can be constructed by choosing the lower endpoint to be the 2.5th and the upper endpoint to be the 97.5th percentile of the posterior distribution. The Bayesian interpretation is that the fixed interval has a 100 (1–α)% probability of containing the random parameter. In the example above, a 95% credible interval for θ would be the 2.5th and 97.5th percentiles of a beta distribution with parameters 7 and 8, which would give an interval of 0.24 to 0.72. Bayesian methods can be applied when multiple parameters are of interest (e.g. multiple regression), in which case a joint prior is specified for the set of parameters and a joint posterior is the outcome.

Specification of a (joint) prior distribution for the parameter(s) of interest is a key part of any Bayesian analysis. One common approach is to specify prior distributions under an assumption of ignorance, so-called non-informative priors. Practically, a non-informative prior distribution arises in a setting where prior information about a parameter is lacking. Some have argued that non-informative priors should always be used as this introduces objectivity into what otherwise appears to be a subjective enterprise. When working with multiple parameters, several non-informative priors are available, and care is needed to ensure the posterior distribution is a valid probability distribution. Carlin and Louis (1996:33–37) and Gelman et al. (1995:52–57) discuss non-informative priors. In some applications, the strength of the data is such that it overwhelms the prior, and the posterior has little dependence on the prior. When the posterior does depend on the prior, the Bayesian would argue that an honest analysis must include discussion of prior choice and recognition that the data do not support a single answer.

Received: 18 May 2001.
Accepted: 6 March 2002.
Associated Editor: Hovey.